DOCUMENT RESUME

ED 264 263 TM 850 748

AUTHOR Lewis, Charles
TITLE Developments in Nonparametric Ability Estimation.
PUB DATE Jul 82
NOTE 19p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Oriented Programs; Estimation (Mathematics); *Item Analysis; *Latent Trait Theory; *Nonparametric Statistics; Psychometrics; *Scaling; Secondary Education; Statistical Studies; *Test Theory
IDENTIFIERS Scholastic Aptitude Test; *Unidimensional Scaling

ABSTRACT

The nonparametric approach to test theory discussed here has its roots in the early work of Guttman, Lazarsfeld, and Meredith; and more recently in the work of Cliff and in Tatsuoka and Tatsuoka. Mokken's extensive treatment of this subject concentrated on defining, constructing, and testing unidimensional scales, based on responses to dichotomous items. His emphasis is on the development of formal statistical methods to extend and support the use of descriptive indices. The analyses described assume the existence of a scale such as might result from one of Mokken's procedures. Based on responses to items forming the scale, inferences are made regarding the abilities of individuals. In this sense, the work described is viewed as an extension of Mokken's developments and includes a discussion of the following: (1) basic assumptions and notations; (2) alternative ordinal assumptions and corresponding definitions of ability; (3) minimal assumption (relative consistency) and analysis; and (4) an example of an analysis. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Developments in Nonparametric Ability Estimation

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Charles Lewis
University of Groningen

ED264263

The nonparametric approach to test theory to be discussed here has its roots in early work of Guttman, Lazarsfeld, and Meredith. References to much of the more recent work in this tradition may be found in Cliff (1979) and in Tatsuoka and Tatsuoka (in press). The most extensive treatment of the subject to date is that of Mokken (1971). Recent introductions to this work are given by Stokman and Van Schuur (1980) and by Mokken and Lewis (1982). Computer programs developed to carry out the relevant analyses are described in The STAP User's Manual (Technisch Centrum FSW, 1980). Mokken (1971) concentrated on defining, constructing, and testing unidimensional scales, based on responses to dichotomous items. Unlike most of the work in this nonparametric tradition, Mokken's emphasis is on the development of formal statistical methods to extend and support the use of descriptive indices.

The analyses to be described here assume the existence of a scale such as might result from one of Mokken's procedures. Based on their responses to items forming the scale, inferences are made regarding the abilities of individuals. In this sense, the work to be described here can be seen as an extension of Mokken's developments. Earlier references are Lewis (1981) and Mokken and Lewis (1982).

Basic Assumptions and Notation

As in classical test theory and virtually all of modern item response theory (IRT), a propensity distribution for observed score, given a particular item and person, is assumed in the following. Since attention will be restricted to dichotomous items, this must be a Bernoulli distribution and will be written as

$$\text{Prob}\,(u_{ia} = 1) = \pi_{ia} \qquad [1]$$

$$\text{Prob}\,(u_{ia} = 0) = 1 - \pi_{ia} \qquad [2]$$

for item i, person a, and observed score u_{ia}. The probability of "success" π_{ia}, which (with 0/1 scoring) is also the classical true score for u_{ia}, will serve as the basic unknown (parameter).

The only other basic assumption which will be made is that given a set of π_{ia}, the corresponding u_{ia} are conditionally independent. This allows their joint propensity distribution to be written as a product of the marginal distri-

TM 850 748

butions and is an assumption common to virtually all developments in modern IRT. (It is, however, stronger than the "uncorrelated errors" assumption of classical test theory.)

As developed in the next section, all additional assumptions will have to do only with the ordering of the π_{ia}. This, indeed, is the reason for using the label "nonparametric" to describe the present treatment. There will be no attempt made here to weigh the advantages and disadvantages of parametric and nonparametric approaches to IRT. The reader who is interested in the "robustness" of various approaches should, however, recall that a technique which never makes a certain assumption is automatically robust to violations of that assumption.

Alternative Ordinal Assumptions and Corresponding Definitions of Ability

Beyond the basic assumptions outlined in the last section, it will be useful to consider ordinal assumptions about the π_{ia} made in previous work, in order to motivate the assumption that will finally be used here. Since the primary emphasis of the present development is on abilities, definitions of ability that are justified by the various assumptions will also be considered.

There are two basic sorts of ordering of the π_{ia} that will be discussed. Let I be a finite set of items which are of interest for measuring a given ability, and let P be the population of persons whose abilities are of potential interest.

First, consider P by I ordering: consistent ordering of persons by items. The idea is that the π_{ia} for each item i should order the persons in the same way. Formally, for items i and j in I and persons a and b in P,

$$\pi_{ia} > \pi_{ib} \text{ implies } \pi_{ja} \geq \pi_{jb}. \qquad [3]$$

(If item i is easier for person a than for person b, then item j should be at least as easy for person a as for person b.)

Second, consider I by P ordering: consistent ordering of items by persons. This is really just a dual of the first type of ordering and states that the π_{ia} for each person a should order the items in the same way. Formally, for items i and j in I and persons a and b in P,

$$\pi_{ia} > \pi_{ja} \text{ implies } \pi_{ib} \geq \pi_{jb}. \qquad [4]$$

(If item i is easier than item j for person a, then i should be at least as easy as j for person b.)

As an illustration of these two types of ordering, a hypothetical set of π_{ia} for a Guttman scale is presented in Table 1 for four items and seven persons. Note first that due to the deterministic nature of Guttman's theory, all π_{ia} are either zero or unity. More importantly, within a Guttman scale, both types of ordering hold, and the rows and columns of Table 1 have been arranged

to reflect this. Thus, for example, Person 5 will (with probability 1) give a positive response to Item 3, while Person 4 will not. P by I ordering implies that Person 5 will do at least as well as Person 4 on the remaining items. (In fact, their response probabilities are identical except for those to Item 3, so the implication holds.) For an example of the second (I by P) ordering, begin with Person 3, who can answer Item 2 but not Item 3. It is also the case that Item 2 is at least as easy as Item 3 for the other six persons, as Equation 4 requires.

Table 1
Hypothetical π_{ia} Values for Four Items and Seven Persons, as an Illustration of the Double Monotony Ordering Assumption for a Guttman Scale

Items in Order	Persons in Order						
	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2	0	0	1	1	1	1	1
3	0	0	0	0	1	1	1
4	0	0	0	0	0	0	1
Sum*	0	1	2	2	3	3	4

*Number Correct

As noted above, a Guttman scale will always satisfy both I by P and P by I ordering. This is also true for Rasch scales. Mokken (1971) used the term double monotony to identify this property in general. Thus, a set of items, I, is doubly monotone with respect to a population of persons, P, when I by P and P by I ordering both hold for all the corresponding probabilities π_{ia}. While double monotony is a highly desirable property, in the sense that it allows strong conclusions to be drawn about the persons and items for which it holds (Mokken, 1971), it is also a highly restrictive assumption that is unlikely to hold exactly for real items and persons. Thus, it is of interest to see what can be done when one or another aspect of double monotony is abandoned.

P by I Ordering Only

As a first step in this direction, suppose that persons may be ordered by items in terms of the π_{ia} (P by I ordering) but that the dual ordering assumption does not necessarily hold. Mokken (1971) referred to this property as monotone homogeneity and considered it the minimum necessary condition for asserting that a set of items measure the same (unidimensional) characteristic for a population of persons. As he showed, P by I ordering allows each person a to be assigned an ordinal latent trait value θ_a and allows nondecreasing item response functions P_i to be defined by

$$\pi_{ia} = P_i(\theta_a) .$$

A hypothetical example of monotone homogeneity is given in Table 2 (in a format similar to that of Table 1). That the persons can be ordered consistently by the items may be seen by noting that the π_{ia} values in <u>each</u> row of the table form a nondecreasing sequence from left to right. (To eliminate decimal points, π_{ia} is multiplied by 10 in the following tables). Noting that Person 2 finds Item 1 easier than Item 2, while the opposite holds for Person 4, establishes that the π_{ia} in Table 2 violate the I by P ordering assumption.

For convenience, θ_a in Table 2 has been identified as the sum over <u>i</u> of π_{ia}. This is simply the true score corresponding to the number-correct observed score for these four items. It reproduces the ordering of the persons given by the individual items, and this fact (among others) has been used by Mokken (1971) to justify the use of the number correct as an indication of ability when working with monotonely homogeneous items. Any strictly increasing transformation of the θ_a in Table 2 would, of course, be equally satisfactory as an ordinal ability measure.

Table 2
Hypothetical π_{ia}(×10) Values, as an Illustration of P by I Ordering (Monotone Homogeneity)

	Persons in Order						
Items	1	2	3	4	5	6	7
1	1	5	6	6	8	9	9
2	1	4	6	7	7	9	9
3	1	2	3	4	8	8	9
4	1	1	4	4	4	4	5
θ_a(Sum)	.4	1.2	1.9	2.1	2.7	3.0	3.2
θ^*_a(Number of $\pi_{ia} \geq .5$)	0	1	2	2	3	3	4

Monotone homogeneity (P by I ordering) is a property shared by most item response models, including those which describe π_{ia} as a 2- or 3-parameter logistic or normal ogive function of ability. By giving up the second aspect of double monotony (namely, I by P ordering), item response functions are allowed to "cross" one another, in which case it is no longer possible to state unambiguously that one item is at least as difficult as another for all persons.

<u>I by P Ordering Only</u>

Double monotony may also be relaxed in the other direction, giving up P by I ordering but maintaining I by P. This case is discussed by Lewis (1981). Here it is clear that the definition of ability based on monotone homogeneity cannot be used (since items need not order persons consistently). If, however, an <u>item mastery level</u> π_m is introduced, a <u>latent number-correct score</u> θ^*_a may be defined as the number of items <u>i</u> in I for which $\pi_{ia} \geq \pi_m$. This alternative defi-

nition of ability does not in itself depend on any ordering assumptions about the π_{ia}. If, however, I by P ordering holds, then

1. All persons with the same latent number-correct score θ_a^* have mastered the same items (at mastery level π_m); and
2. $\theta_a^* > \theta_b^*$ implies that the items mastered by person b form a subset of those mastered by person a.

These properties of latent number correct are a direct stochastic generalization of those which hold for observed number correct in the case of a Guttman scale. They allow θ_a^* to be used to measure ability on a labeled ordinal scale. Property 1 states that values of θ_a^* may be labeled by the item(s) that have been mastered and Property 2 gives a specific meaning to the ordering of θ_a^*, again in terms of the items that have been mastered.

As an illustration, consider the hypothetical π_{ia} values given in Table 3. The items are ordered consistently by the persons, as can be seen from the fact that each column of π_{ia} in the table gives a nonincreasing sequence from top to bottom. That the persons are not ordered consistently by the items is shown by the fact that Item 1 is easier for Person 2 than for Person 1, while the opposite is true for Item 2. A mastery level of .5 has been chosen for determining θ_a^*, whose values range from 0 for Person 1 (reflecting that all π_{i1} are less than .5) to 4 for Person 7 (all π_{i7} being greater than or equal to .5). The fact that θ_a^* is 3 for Persons 5 and 6 implies that both have mastered Items 1, 2, and 3, but not 4 (at level .5). It also implies that they have mastered all items mastered by those persons (1 through 4) for which θ_a^* is less than 3.

Models where I by P ordering is assumed (but not double monotony) are unusual within IRT. An example is the approach based on person characteristic curves, which was discussed by Lumsden (1978). In this approach it is assumed that the π_{ia} may be written as nonincreasing functions (one for each person) of a latent item difficulty parameter. The "nonincreasing" property guarantees I by P ordering. However, if two of the functions cross, then P by I ordering is violated, since an item "below" the crossing would order the two persons one way and an item above would do the reverse.

Table 3
Hypothetical π_{ia} (×10) Values, as an Illustration of I by P Ordering

Items in Order	Persons						
	1	2	3	4	5	6	7
1	4	5	6	7	8	8	9
2	3	2	4	6	8	8	9
3	1	2	4	3	7	6	8
4	1	1	2	3	4	3	5
θ_a^*	0	1	1	2	3	3	4

Comparing θ_a and θ_a^*

Returning to the definitions of ability, θ_a and θ_a^*, Mokken and Lewis (1982) showed that they have a natural relation in the case of double monotony. Briefly summarized, θ_a^* defines a partitioning of θ_a into ordered equivalence classes:

$$\Theta_j = \{\theta_a | \theta_a^* = j\} , \qquad [6]$$

where $j = 0(1)n$ in the case of n items. (These are ordered in the sense that $i > j$ implies that every element in Θ_i exceeds every element in Θ_j.) Thus, θ_a^* could be thought of as a discretizing and labeling of θ_a, when double monotony holds for the π_{ia}.

Mokken (1982), has pointed out that the same relationship between θ_a and θ_a^* also exists in the general case of monotone homogeneity (thus, when I by P ordering need not hold). Although the items do not have an unambiguous order in this case, they may still be ordered relative to a given mastery level π_m. This may be illustrated by returning to the example of monotone homogeneity given in Table 2. If .5 is adopted as a mastery level, the items are ordered consistently with their numbering. Thus, Item 1 is the easiest, since everyone except Person 1 has mastered it, and Item 4 is the most difficult, since only Person 7 has mastered it. (That this ordering depends on the choice of mastery level may be seen by observing that if $\pi_m = .4$, Item 3 replaces Item 4 as the most difficult.) The last row of Table 2 gives values of θ_a^* based on $\pi_m = .5$, and a comparison of these values with those of θ_a in the row above illustrates the relationship between the two. Note that although some of the inequalities among the θ_a values become equalities among θ_a^* (this is what was meant by discretizing), none of the inequalities is reversed. In other words, whenever monotone homogeneity holds, then θ_a^* may be thought of as measuring the same thing as θ_a, but doing it more coarsely. On the other hand, the labeling provided by θ_a^* may be considered as a useful return for this investment. Thus, returning to Table 2, it may be inferred from θ_a that Persons 3 and 4 have higher ability than Persons 1 and 2 but lower ability than Persons 5, 6, and 7. From θ_a^*, however, it is also known that Persons 3 and 4 have mastered Items 1 and 2, but not Items 3 and 4 (at the .5 level).

A "Minimal" Assumption (Relative Consistency) and Analysis

With the above extended discussion as background, the ordering assumption that will actually be used in the following may be introduced. The idea is to work with an assumption that allows item labeling of the latent number-correct score θ_a^* but is otherwise as weak as possible. To achieve this goal, neither P by I nor I by P ordering is necessary. Only consistent ordering of the π_{ia} relative to the item mastery level π_m used to define θ_a^* need be considered. The actual assumption will be referred to as relative consistency.

Specifically, begin with any pair of persons in P and any pair of items in I. If each of the persons can master just one of the two items (at level π_m), the assumption is that this is the same item for both persons. Phrased this way, the persons may be thought of as consistently ordering the items relative to π_m, a generalization of I by P ordering.

This asymmetry is actually misleading. The assumption may be equivalently stated as follows. If each of the two items can be mastered (at level π_m) by just one of the two persons, then this must be the same person for both items. Thus, the items must consistently order the persons relative to π_m, generalizing P by I ordering.

To see the equivalence of these two statements of the assumption of relative consistency, consider a formal negative expression for both: There are no items i and j in I and persons a and b in P such that the following four relations simultaneously hold:

$$\pi_{ia} \geq \pi_m \ , \quad \pi_{ib} < \pi_m \ ,$$

$$\pi_{ja} < \pi_m \ , \quad \pi_{jb} \geq \pi_m \ . \qquad [7]$$

Read by rows, Equation 7 can be seen to refer to an inconsistent ordering of the persons a and b by the items i and j. Read by columns, it refers to an inconsistent ordering of the items by the persons.

For the purposes of what follows, the importance of assuming that the pattern in Equation 7 does not occur (for the π_{ia} associated with sets P and I of interest) is to guarantee that the earlier mentioned properties--same θ_a^* score, implying same items mastered, and lower θ_a^* score, implying subset of items mastered--hold, so that θ_a^* forms a labeled ordinal scale.

To illustrate this new assumption, hypothetical π_{ia} values are given in Table 4. Using a mastery level of .5, both items and persons may be consistently ordered. (This has been done in the table.) To make verification of the two properties of θ_a^* easier, a "staircase" has been inserted in the table, separating mastery from nonmastery levels of the π_{ia}.

When π_m is changed to .4, the π_{ia} associated with the first two items and the first two persons exhibit the inadmissible pattern in Equation 7 . Thus, Person 1 can only master Item 2 and Person 2 can only master Item 1 (at level .4). This shows that relative consistency does not hold for the π_{ia} in Table 3 when $\pi_m = .4$. It also shows that neither I by P nor P by I ordering holds in this example, since the presence of pattern Equation 7 for any π_m is inconsistent with both these assumptions.

Table 4
Hypothetical π_{ia} (×10) Values, as an Illustration of Relative Consistency ($\pi_m = .5$)

Items in Order	Persons in Order						
	1	2	3	4	5	6	7
1	3	5	7	5	9	6	9
2	4	2	4	6	8	8	7
3	4	2	3	3	7	6	8
4	1	1	2	4	4	3	5
θ_a^*	0	1	1	2	3	3	4

Statistical Analysis of an Individual Response Pattern

If the responses of individual a to Items 1 through n are denoted by

$$\underset{\sim}{u}_a' = (u_{1a},\ldots,u_{na}) \qquad [8]$$

and the corresponding probabilities of positive response by

$$\underset{\sim}{\pi}_a' = (\pi_{1a},\ldots,\pi_{na}), \qquad [9]$$

then the basic assumption of conditional independence of the u_{ia} implies

$$p(\underset{\sim}{u}_a|\underset{\sim}{\pi}_a) = \prod_{i=1}^{n} \pi_{ia}^{u_{ia}} (1 - \pi_{ia})^{1-u_{ia}} . \qquad [10]$$

For reasons discussed in detail by Lewis (1981) and mentioned by Mokken and Lewis (1982), a Bayesian aproach to making inferences about θ_a^* will be adopted here. Thus, a prior distribution for $\underset{\sim}{\pi}_a$ must be specified to describe available prior knowledge. In the above mentioned references, critical use was made of I by P ordering in selecting the prior. Here, priors will be considered which are based on the relative consistency assumption. More concretely, not only will relative consistency be assumed, but it will also be assumed that the ordering of the items relative to the chosen mastery level (π_m) is known and that the items have been numbered (from 1 to n) to reflect this ordering. Thus, for items i and j in I, $i < j$ implies that for all persons a in P, if $\pi_{ja} \geq \pi_m$, then $\pi_{ia} \geq \pi_m$. (In other words, anyone who can master item j can also master item i.)

Although the π_{ia} will not, in general, have a functional dependence on θ_a^*, it may be convenient to express prior knowledge about $\underset{\sim}{\pi}_a$ conditional on θ_a^*. Thus, the joint prior may be decomposed as

$$p(\underset{\sim}{\pi}_a, \theta_a^*) = p(\underset{\sim}{\pi}_a|\theta_a^*)p(\theta_a^*). \qquad [11]$$

Relative consistency implies that the π_{ia} cannot, in general, be stochastically independent of one another. Given θ_a^*, however, such an assumption is not ruled out and is, in fact, very convenient. Thus, it will be assumed that prior knowledge is such that the conditional distribution of $\underset{\sim}{\pi}_a$ given θ_a^* may be written as

$$p(\underset{\sim}{\pi}_a|\theta_a^*) = \prod_i p(\pi_{ia}|\theta_a^*). \qquad [12]$$

This assumption will be referred to as <u>second-order conditional independence</u>.

Note that the limits within which the density $p(\pi_{ia}|\theta_a^*)$ may be positive vary with the value of θ_a^*. Thus, if $\theta_a^* = j$,

$$\pi_{ia} \geq \pi_m \quad \text{for } i \leq j \quad \text{and } \pi_{ia} < \pi_m \quad \text{for } i > j. \qquad [13]$$

In other words, if person a has a latent number-correct score of j, then the <u>first j</u> items must have been mastered (at level π_m) and the <u>last</u> n-j items <u>not</u> mastered.

Continuing further with the analysis, Bayes' theorem may be used with Equations 10, 11, and 12 to obtain the joint posterior distribution for $\underset{\sim}{\pi}_a$ and θ_a^*, given $\underset{\sim}{u}_a$:

$$p(\underset{\sim}{\pi}_a, \theta_a^*|\underset{\sim}{u}_a) \propto p(\theta_a^*) \prod_i \left[\pi_{ia}^{u_{ia}} (1-\pi_{ia})^{1-u_{ia}} p(\pi_{ia}|\theta_a^*) \right]. \qquad [14]$$

For most applications, primary interest will center on the marginal posterior for θ_a^*. That is to say, inferences about latent number correct will typically be more important than direct statements about the individual π_{ia}.

The marginal distribution for θ_a^* may be obtained by integrating the joint posterior Equation 14 with respect to $\underset{\sim}{\pi}_a$, separately for each value of θ_a^*. Due to the first- and second-order independence assumptions of Equations 10 and 12, the desired multiple integral may be expressed as a product of single integrals. These will now be examined in more detail.

If $u_{ia} = 1$, the corresponding integral for π_{ia} will have the form

$$\int \pi_{ia} \, p(\pi_{ia}|\theta_a^*) \, d\pi_{ia} = E(\pi_{ia}|\theta_a^*) , \qquad [15]$$

the prior conditional expectation of π_{ia} given θ_a^*, which may be denoted by

$$\bar{P}_i(\theta_a^*) = E(\pi_{ia}|\theta_a^*) . \qquad [16]$$

Similarly, if $u_{ia} = 0$, the integral becomes

$$\int (1-\pi_{ia}) p(\pi_{ia}|\theta_a^*) d\pi_{ia} = 1 - \bar{P}_i(\theta_a^*) \qquad [17]$$

in the new notation. Combining the results of Equations 15 and 17 gives the following expression for the marginal posterior distribution for θ_a^*:

$$p(\theta_a^*|\underset{\sim}{u}_a) \propto p(\theta_a^*) \prod_i [\bar{P}_i(\theta_a^*)]^{u_{ia}} [1-\bar{P}_i(\theta_a^*)]^{1-u_{ia}} . \qquad [18]$$

Equation 18 may be seen as the most important single result in the present development and, as such, should be discussed in some detail. The most remarkable aspect of Equation 18, especially considering the minimal assumptions needed to obtain it, is its simplicity. Formally, it is identical to the result which is obtained beginning with the assumption that the π_{ia} are known functions of a postulated latent trait, using conditional independence of the u_{ia}, and introducing a prior distribution for the latent trait (see, for example, Birnbaum, 1969). In fact, in the present development, the latent trait (θ_a^*) is constructed in terms of the π_{ia}, making use only of the relative consistency assumption with regard to the latter.

Turning to the problem of specifying the prior density of Equation 11 for θ_a^* and $\underset{\sim}{\pi}_a$, Equation 18 makes it clear that only the marginal prior for θ_a^* and the conditional means $\bar{P}_i(\theta_a^*)$ need be considered. The precise form of the $p(\pi_{ia}|\theta_a^*)$ found in Equation 12 is irrelevant to the goal of making inferences about θ_a^*.

On the other hand, in specifying the conditional means, it should be recognized that the choices have implications for the corresponding conditional variances. In particular, given the restricted ranges of the π_{ia} as indicated in Equation 13, plus the inherent restriction to the interval (0,1), specifications of $\bar{P}_i(\theta_a^*)$ close to 0, π_m, or 1 imply that $\mathrm{Var}(\pi_{ia}|\theta_a^*)$ must be very small. If uncertainty regarding the true value of π_{ia} for a given value of θ_a^* is present in any substantial degree, it then follows that such "boundary" specifications should be avoided.

One concrete way of making appropriate specifications for $\bar{P}_i(\theta_a^*)$ will now be discussed. Suppose that a density from the beta family is chosen for $p(\pi_{ia}|\theta_a^*)$ and then truncated above or below π_m to conform with the restrictions in Equation 13. This choice has the advantages of providing a (conditional) prior for $\underset{\sim}{\pi}_a$, which is a natural conjugate to the likelihood of Equation 10, and (except for the truncation) of being familiar to users of elementary Bayesian techniques.

As described, for instance, in Novick and Jackson (1974), the choice may be completely specified by identifying the sum of the beta parameters ($\alpha + \beta$) with a hypothetical prior sample size (larger values reflecting more precise prior knowledge) and the mode of the density [$(\alpha - 1)/(\alpha + \beta - 2)$ for $\alpha, \beta > 1$] with the a priori most likely value for π_{ia}, given θ_a^*. The conditional means $\bar{P}(\theta_a^*)$ of the truncated densities may then be computed in terms of the untruncated mean $(\alpha/(\alpha + \beta))$ and incomplete beta functions as follows:

$$\bar{P}_i(\theta_a^*) = \begin{cases} \left(\frac{\alpha}{\alpha+\beta}\right)[1-I_{\pi_m}(\alpha+1,\beta)]/[1-I_{\pi_m}(\alpha,\beta)] & \text{for } i \leq \theta_a^*, \text{ and} \\ \left(\frac{\alpha}{\alpha+\beta}\right) I_{\pi_m}(\alpha+1,\beta)/I_{\pi_m}(\alpha,\beta) & \text{for } i > \theta_a^*. \end{cases} \quad [19]$$

Here α and β are the parameters of the chosen beta density for π_{ia} given θ_a^* (and, thus, should formally have subscripts $\underline{i}$, $\underline{a}$, and θ_a^*).

The procedure just described provides a formal mechanism for "adjusting" prior estimates for π_{ia} away from the boundaries 0, π_m, and 1 to a degree which is controlled by the choice of hypothetical prior sample size ($\alpha + \beta$), with more precise prior knowledge leading to less adjustment. As mentioned earlier, the precise form of the truncated beta has no further consequences for inferences about θ_a^* and it should be recognized that there will be a range of densities having the same (or almost the same) modes and means as those obtained from any given choice of truncated betas.

Finally, a word should be said about the specification of $p(\theta_a^*)$. This will be a discrete distribution and it would seem wisest not to restrict its form in practice to that of any particular family, such as the binomial. It will often be the case that some information about the population distribution of ability in P is available from analyses carried out during the development of the test. Unless extra information is available regarding the specific person $\underline{a}$, this population information will provide a good basis for specifying $p(\theta_a^*)$. In such a case, it may be useful to remember that Prob$(\theta_a^* \geq j)$ should ideally correspond to the proportion of persons in P who can master item $\underline{j}$ at level π_m.

Example of an Analysis

As noted at the beginning of this paper, the present treatment of abilities is most directly compatible with the work of Mokken (1971). In this spirit, Mokken and Lewis (1982) discuss an example in which the procedures proposed by the first author are applied for test development and evaluation, and those by the second author for the test administration phase (to make inferences about individual abilities). Suppose, however, that information regarding an existing test is only available in terms of a parametric item response model. What are the possibilities for analyzing individual responses to items in such a test using the approach described in the previous section? In what follows, this question will be considered in terms of a simple example.

Lord (1968) applied the 3-parameter logistic item response model,

$$\pi_{ia} = P_i(\theta_a) = \{1+\exp[-1.7a_i(\theta_a - b_i)]\}^{-1}(1-c_i) + c_i \ , \quad [20]$$

to analyze the responses of nearly 3,000 persons to items of the Verbal Scholastic Aptitude Test. Most importantly for present purposes, he provided a table of estimated item parameters (a_i, b_i, and c_i) and of selected percentiles

of the distribution of estimated θ_a values. While questions may be legitimately raised about various aspects of this analysis, as Lord himself did (for example, does the model provide a reasonable fit to the data?), the results are undoubtedly an important source of prior information for anyone wishing to work further with these items. In the present context, the problem becomes one of translating this information into the components required in Equation 18 to carry out an analysis of individual responses.

Remembering that θ_a^* may be thought of as dividing the θ_a scale into intervals, it would be convenient if the percentiles for θ_a given by Lord (1968) could be used to define boundaries for these intervals. In fact, the boundaries must be determined by individual items, relative to a given mastery level. For this example, an item mastery level $\pi_m = .6$ was chosen. This may be thought of as a correction of the "natural" value of .5 for "guessing," given that these are five-alternative multiple-choice items. Using Lord's tabled item parameters, items were then sought whose estimated response functions attained .6 for values of θ_a close to those given for reported percentiles. Since Lord provided parameters for 80 items of widely varying difficulty, this search ended successfully for six of the seven percentiles given (the exception being the .0025th). The results are summarized in Table 5.

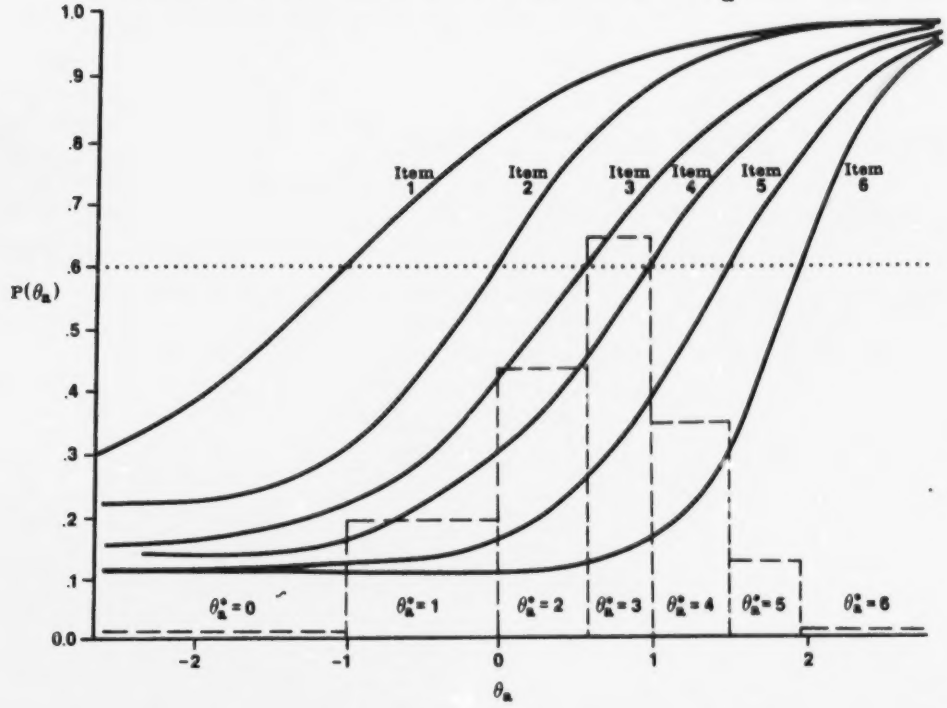
Table 5
Sample Items Chosen from Lord (1968)

Item No.	SAT No.	Reported Item Parameters a_i	b_i	c_i	θ_a Giving $P_i(\theta_a)=.6$	Reported Percentile
1	52	0.7	-1.0	.20	-1.00	.02
2	62	1.1	0.0	.20	0.00	.21
3	56	0.9	0.5	.14	0.59	.50
4	79	1.0	0.9	.13	0.99	.76
5	15	1.2	1.4	.11	1.50	.935
6	66	1.9	1.9	.11	1.96	.99

In the analysis which follows, only the six items described in Table 5 will be considered. This greatly simplifies presentation of results and makes it possible, as already indicated, to use the reported percentiles in selecting a prior distribution for θ_a^*. The estimated response functions for these items, obtained by substituting the parameter values from Table 5 in Equation 20, are shown in Figure 1, as is the relative frequency distribution for the estimated θ_a values, displayed as a histogram and based on the percentiles from Table 5.

Turning now specifically to the choice of priors, Table 6 gives the prior distribution for θ_a^* using the reported percentiles. Such a specification implies that the abilities of persons whose responses will be analyzed may be thought of as exchangeable with those for the group analyzed by Lord and that the abilities estimated in Equation 20 are thought to have a relationship to the

Figure 1
Item Response Functions and Distribution of θ_a from Table 5



θ_a^* defined by the π_{ia} for these six items and item mastery level of .6. Both of these implications refer to subjective evaluations rather than to objective facts regarding the π_{ia}. The requirement, thus, is that the implications be "reasonable," not that they be "true." In fact, as Lord (1968, p. 1018) has explained, for technical reasons, persons with extreme abilities (and especially low abilities) were over-represented in his sample. Thus, the prior of Table 6 is almost certainly too "low" and too "broad" for use with "typical" SAT examinees. This will be ignored in what follows.

Table 6
Prior Distribution for Latent Number Correct (θ_a^*), Based on Data in Table 5, with $p(\theta_a) \times 100$

Variable	Items						
	1	2	3	4	5	6	
θ_a^*	0	1	2	3	4	5	6
Prior	2	19	29	26	17.5	5.5	1

The prior distribution for θ_a^* having been chosen, attention may be given to selecting conditional priors for the π_{ia} given θ_a^*. Following the approach of

the previous section, truncated beta distributions will be considered for this purpose.

Most likely values (prior conditional modes) for the π_{ia} may be adopted from the estimated item response functions pictured in Figure 1. For simplicity, each function was evaluated at the midpoint of each of the finite θ_a intervals, corresponding to θ_a^* values 1 through 5. For the infinite intervals corresponding to θ_a^* of 0 and 6, the functions were evaluated at points with a distance equal to the average of the finite interval widths below the lowest and above the highest finite boundary, respectively.

The adoption of the actual sample size on which the estimates were based (2,862) as the hypothetical prior sample size for the betas might be considered. This would lead to conditional prior means for the π_{ia} virtually identical to the most likely values and would reflect complete confidence in the estimated functions obtained by Lord as descriptions of the π_{ia}. Such confidence is almost certainly not justified, and it may be argued that modal values close to 0, .6, and 1 required considerable adjustment to reflect uncertainty regarding the π_{ia} for an individual with ability θ_a^*. After some experimentation, a hypothetical prior sample size of 10 was adopted for all the conditional priors, and the required $\overline{P}_i(\theta_a^*)$ computed.

Figure 2 illustrates the process for Item 3 and $\theta_a^* = 2$. The truncated beta density for π_{3a} has its mode at the value given by the item response function $P_3(\theta_a)$ for θ_a at the midpoint of the interval corresponding to $\theta_a^* = 2$. Because of the truncation above .6, the conditional mean is somewhat lower than the mode, as indicated in the figure. The conditional means $\overline{P}_3(\theta_a^*)$ for the remaining values of θ_a^* are displayed as a step function in Figure 2 and indicate the extent of adjustment away from the boundaries obtained with a prior sample size of 10.

The full results are given in Table 7, where a "staircase" has been added to separate means for the items which, by definition, are mastered at each ability level from the rest. Thus, values above the staircase are all greater than the mastery level of .6 and those below all less than .6. A few extreme examples of adjustment from mode to mean are: .99 → .90 for Items 1 and 2, Ability 6; .11 → .19 for Items 5 and 6, Ability 0; .66 → .73 for Item 3, Ability 3; and .53 → .44 for Item 4, Ability 3.

Tables 6 and 7 provide all the prior information necessary to carry out via Equation 18 posterior analyses for individual abilities, based on responses to the six items of Table 5. Of course, it may reasonably be asked how much the prior distribution of Table 6 could possibly be modified by the information contained in only six responses. To shed some light on this question, consider first the posterior distributions based on two (hypothetical) extreme response records: all correct and none correct. These are shown in Table 8.

Figure 2
Adjustment from $P_3(\theta_a)$ to $\overline{P}_3(\theta_a^*)$, with $p(\pi_{3a}|\theta_a^* = 2)$ Shown Explicitly

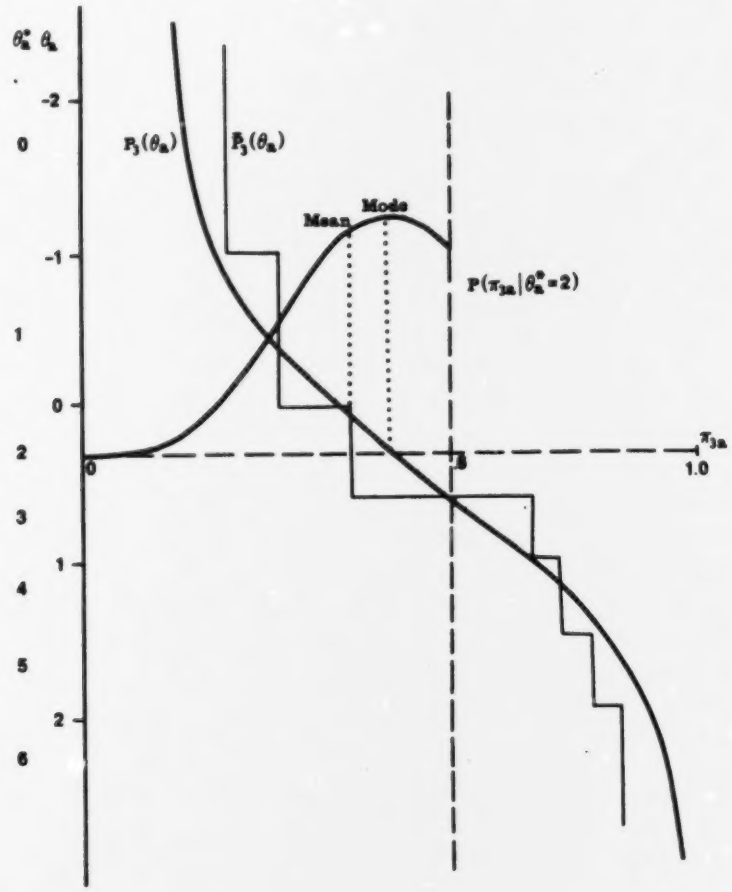


Table 7
Conditional Prior Means for π_{ia} Given θ_a^* for Hypothetical Example, with $\overline{P}_i(\theta_a^*) \times 100$

Items	θ^* 0	1	2	3	4	5	6
1	42	75	81	85	87	88	90
2	28	40	74	81	85	88	90
3	24	32	43	73	78	83	88
4	21	26	36	44	74	80	87
5	19	20	25	33	43	74	85
6	19	19	19	21	26	40	84

The prior distribution for ability was chosen so that the probability assigned to $\theta_a^* \geq 3$ was just .5. It will be remembered that this may be interpreted as the a priori probability that person a can master Item 3 at the .6 level. How much has this been modified in the two distributions given in Table 8? The a posteriori probability that a person answering all six items correctly can master Item 3 is .945, the sum of the probabilities for $\theta_a^* = 3, 4, 5$, and 6. For a person answering none of the six items correctly, the probability that Item 3 can be mastered is only .04.

Table 8
Posterior Distributions for θ_a^* Based on Two Hypothetical Extreme Response Records ($\underset{\sim}{u}_a$), with $p(\theta_a^* | \underset{\sim}{u}_a) \times 100$

Record and Variable	Items						
	1	2	3	4	5	6	
θ_a^*	0	1	2	3	4	5	6
Record 1							
Responses	1	1	1	1	1	1	
Posterior	0	0.5	5	15	31	32	16.5
Record 2							
Responses	0	0	0	0	0	0	
Posterior	20	57.5	18.5	3.5	0.5	0	0

Thus, at least for these two extreme cases, it must be concluded that the responses have a substantial effect on the prior. That the prior also plays an important role in this example may be seen by noting that the posterior probability that a person answering all six items correctly can actually master all these items ($\theta_a^* = 6$, all $\pi_{ia} \geq .6$) is only .165. Similarly, the probability is only .20 that a person with none correct can actually master none ($\theta_a^* = 0$, all $\pi_{ia} < .6$).

As a final illustration of the interaction between prior and data in this example, sensitivity to response pattern will be considered. Once again, two hypothetical response records are analyzed, and the results are shown in Table 9. Each record consists of three correct and three incorrect responses. For simplicity, the first will be referred to as a "perfect" pattern and the second as having a single "reversal."

The perfect record suggests that the first three items can be mastered and, indeed, the highest posterior probability based on this record is assigned to $\theta_a^* = 3$. That the second highest probability is for an ability of 2, rather than 4, is a reflection of the prior, whose mode was 2 (see Table 6). What does the single reversal record suggest? Abilities of 2 and 4 seem more plausible than 3, since the latter implies two reversals rather than one. Comparing the resultant posterior distribution with the previous one, it may be observed that the probabilities for $\theta_a^* = 2$ and 4 are larger, while that for $\theta_a^* = 3$ is smaller.

Table 9
Posterior Distributions for θ_a^* Based on
Two Hypothetical Response Patterns, with $p(\theta_a | \underset{\sim}{u}_a) \times 100$

Record and Variable	Items						
	1	2	3	4	5	6	
θ_a^*	0	1	2	3	4	5	6
Record 1							
Responses	1	1	1	0	0	0	
Posterior	0.5	10	33	43	12.5	1	0
Record 2							
Responses	1	1	0	1	0	0	
Posterior	0.5	13	43.5	23	18	2	0

That the posterior distribution for ability based on a single reversal is unimodal rather than bimodal, and that the mode is 2 rather than 4, may be attributed to the influence of the prior.

Turning to the characteristic discussed for the extreme response records, namely, the posterior probability that Item 3 can be mastered, the perfect record results in a value of .565, while the single reversal leads to a probability of only .43. Thus, it may be concluded that within the present approach even small variations in response pattern can have a noticeable influence on inferences about abilities.

REFERENCES

Birnbaum, A. Statistical theory for logistic mental test models with a prior distribution of ability. Journal of Mathematical Psychology, 1969, 6, 258-276.

Cliff, N. Test theory without true scores? Psychometrika, 1979, 44, 373-393.

Lewis, C. Estimating abilities: Inferences for random variables. Kwantitatieve Methoden, 1981, 2, 17-34.

Lord, F. M. An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lumsden, J. Tests are perfectly reliable. British Journal of Mathematical and Statistical Psychology, 1978, 31, 19-26.

Mokken, R. J. A theory and procedure of scale analysis with applications in political research. New York: de Gruyter/Berlin: Mouton, 1971.

Mokken, R. J. Personal communication, 1982.

Mokken, R. J., & Lewis, C. A nonparametric approach to the analysis of dichotomous item responses. Applied Psychological Measurement, 1982, 6, 417-430.

Novick, M. R., & Jackson, P. H. Statistical methods for educational and psychological research. New York: McGraw-Hill, 1974.

Stokman, F. N., Van Schuur, W. H. Basic scaling. Quality and Quantity, 1980, 14, 5-30.

Tatsuoka, K. K., & Tatsuoka, M. M. Spotting erroneous rules of operation by the individual consistency index. Journal of Educational Psychology, in press.

Technisch Centrum FSW. STAP user's manual (Vol. 4. Stochastic cumulative scaling: Mokken scale, Mokken test). Amsterdam: University of Amsterdam, 1980.